DOCUMENT RESUME

ED 185 088 **TM 800 087**

AUTHOR	Buck, Lawrence S.
TITLE	Is the Federal Government Jumping on the Criterion-Referenced Testing Bandwagon?
PUB DATE	Oct 79
NOTE	13p.: Paper presented at the Annual Meeting of the Northeastern Educational Research Association (Ellenville, NY, October 24-26, 1979).
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Criterion Referenced Tests; *Federal Government; *Government Employees; Job Placement; *Job Training; Military Personnel; *Military Training; Personnel Evaluation; Personnel Selection; Testing Problems; *Testing Programs; Training Objectives
IDENTIFIERS	Civil Service Commission; Equal Employment Opportunity Commission; Internal Revenue Service; Social Security Administration

ABSTRACT

The increasing use of criterion referenced testing (CRT) among the various branches of the federal government is described. The requirements of the merit system have tended to promote the use of norm referenced tests except for uses such as pass/fail performance tests. The two areas in which criterion-referenced tests have been most useful are training and occupational licensure. The military services have lead the way in using criterion-referenced tests in their training programs in spite of their high initial costs because the CRT's are effective screening devices and diagnostic aids and for evaluation of instruction. The Internal Revenue Service has established a centralized testing program capable of providing feedback to course developers, program managers, and instructors. It's manual item banks are being converted to computerized banks. The Social Security Administration has developed an extensive programmed instruction system based on criterion-referenced testing. The Civil Service Commission's Bureau of Training is considering the use of criterion-referenced testing to evaluate training programs. Properly developed criterion-referenced tests should meet the requirements of the Uniform Guidelines on Employee Selection Procedures. (CTM)

****
* Reproductions supplied by EDPS are the best that can be made *
* from the original document. *
****

ED185088

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Is the Federal Government Jumping on the Criterion-Referenced Testing Bandwagon?

Lawrence S. Buck
United States Office of Personnel Management
Washington, D.C. 20415

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
John D. Kraft

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at
the
Northeastern Educational Research Association Convocation
Ellenville, New York
October 1979

TM800 087

Abstract

Is the Federal Government Jumping On the
Criterion-Referenced Testing Bandwagon?

Lawrence S. Buck
United States Office of Personnel Management

In the decade of the 70's, the criterion-referenced testing movement has experienced a tremendous growth in the private and academic sectors. While criterion-referenced testing principles seem emminently well suited to the testing needs of the Federal government, the movement towards the use of criterion-referenced tests (CRTs) has, until recently, been considerably more subdued. One reason for this apparent lag has been a tendency to stay with the more time-tested and traditional norm-referenced testing approach. Merit system requirements for evaluating the relative fitness of applicants and Veterans Preference Act requirements for ranking of applicants on the basis of examination scores have also been limiting factors in the governmental use of CRTs.

Despite the roadblocks, the Federal government has seen an upsurge in the last couple of years in the use of CRTs. This report discusses some of the ways in which CRTs have been utilized in the government during this decade. Perhaps the most far-reaching use of criterion-referenced testing principles in the government is evolving at present. The impetus for this present movement is the Civil Service Reform Act of 1978 which has mandated the implementation of performance appraisal systems based on criterion-referenced testing principles.

The decade of the 70's has ushered in a tremendous growth in popularity for the criterion-referenced testing movement in the private and academic sectors. During this period, professional journals have been replete with criterion-referenced testing articles, educational and psychological conventions have allotted significant portions of their programs to the subject of criterion-referenced tests (CRTs), and some State legislatures have mandated the use of CRTs in their educational systems. While criterion-referenced testing principles seem emminently well-suited to the testing needs of the Federal government, the use of CRTs in the government has proceeded much more slowly, especially when compared with their use in the private and academic sectors. This is not to say that CRTs are nonexistent in the government, since they have been utilized with great success in many areas, but rather that their potential has barely been scratched.

There are a number of reasons for this apparent lag, not the least of which has been a tendency to stay with the more time-tested and traditional norm-referenced testing approach. Testing in the Federal government has had a long and distinguished history steeped in the tradition of norm-referenced testing methodologies and change does not always come easily. Merit system requirements for evaluating the "relative" fitness of applicants on the basis of examination scores, where examinations are used in the selection process, and the Veterans Preference Act which specifies that examinees be rank ordered according to test score and that Veterans' scores be augmented by five or ten points have been limiting factors. Criterion-referenced tests are not purposefully constructed to limit the number of persons who can qualify for a position or to adapt themselves readily to a ranking scheme and thus are not compatible with merit system and Veterans Preference requirements.

Other problems in the use of CRTs in the government have centered around assessing the validity and reliability of CRTs, standardization of the testing process, and the potential prohibitive costs of developing CRTs.

Despite these limitations, tests which would classify as CRTs have long been in use in the Federal government. Such tests have generally appeared in the form of performance tests graded on a pass/fail basis. The most obvious example being typing performance tests. It is probable that the use of CRTs for selection purposes will increase in the government in the future. The inherent job relatedness and validity of CRTs make them viable selection instruments. Criterion-referenced tests should be appealing to competitors because of their assessment of actual job tasks and because of their face validity. Criterion-referenced tests should also be appealing to employers because they are constructed to yield specific information in terms of a person's performance relative to an established job-related criterion.

There are two areas in which CRTs are particularly applicable and in which they have received considerable attention in the government during this decade--training and occupational licensure. A criterion-referenced testing approach lends itself well both to the specifications of training objectives and to the evaluation and assessment of training objectives and programs. Because of the ability of CRTs to evaluate the examinees' mastery (relative to specified performance standards) for a particular subject-matter area, CRTs are especially appropriate for measurement in the field of occupational licensing. Licensing agencies have the responsibility of identifying those test scores which best distinguish the occupationally competent from those who are not ready for occupational practice. Criterion-referenced tests are well-suited to this end.

In the early part of the decade, the military services led the way as advocates and users of CRTs. They turned to criterion-referenced testing methodologies to improve their training programs and to assess more effectively the abilities and competencies of their personnel. The Experimental Volunteer Army Training Program (EVATP), designed to standardize EVATP instruction, review, and testing, relied extensively on CRTs. Performance tests covering a wide variety of content areas were developed and distributed to training instructors. These tests were used by drill instructors for review or remediation while testing personnel used these tests for general subject, comprehensive performance, and military occupational specialty (MOS) examinations. The performance tests also provided the basis for the EVATP Quality Control System. This system was designed to check on skill acquisition and maintenance during the training process. Although considerable difficulty was encountered in changing the role of the instructors from that required for the traditional well-established instructional system to that required for the new criterion-referenced system, evaluation of the new system revealed clear-cut superiority over the conventional instructional system.

As the decade of the 70's progressed, CRTs received widespread recognition, acceptance, and use within the various military services and in joint efforts among the military services. As traditional barriers and resistances were worn down, the criterion-referenced testing movement was slowed only by the high initial cost of developing criterion-referenced performance tests. The primary use of CRTs in the military services has been in the realm of training where these tests have been called on to perform the following

functions:

1. Screening devices--to assess whether or not an individual possesses required entry behaviors before starting a course and to assess if the individual already knows how to perform some of the tasks;

2. Diagnostic aids--to show what objectives of a course an individual is weak on and/or what particular steps of a procedure the individual is unable to perform;

3. Evaluation of instruction--has an instructional system taught what it is supposed to teach.

Swezey, Pearlstein, and Ton (1974) surveyed a number of Army installations to assess the applications of criterion-referenced testing techniques in the military. The issues investigated included among other relevant issues: how are CRTs developed for Army application; the extent of criterion-referenced testing in the Army; problems encountered in the development and use of CRTs; and attitudes of Army testing personnel toward the development and use of CRTs. The results of the survey indicated that the most common uses for criterion-referenced testing results are: evaluation of trainee performance, improving training, and diagnostic purposes. Nearly three-quarters of those surveyed indicated that the expense of developing and using CRTs is the biggest obstacle to the use of CRTs in the Army. However, many indicated that the increased expense is a short-term factor and that in the long run criterion-referenced testing is less expensive than norm-referenced testing. This is due to the improvements in the training courses facilitated by criterion-referenced testing principles and the improvement in the skills of the trainees and in the methods used to assess the trainees' skills acquisition. The interviewees were generally in favor of the Army trend toward criterion-referenced testing and it was apparent that CRT use is increasing in the Army.

In the latter part of the decade, the Internal Revenue Service (IRS) made a wholesale commitment to the use of CRTs throughout their training system. The IRS has established a centralized testing program designed to establish a centralized test bank for each basic recruit course and to provide test results or hard data that can be used in evaluating the training operation. The centralized testing program is capable of providing:

1. Feedback to course developers and program managers for improvement of course materials;

2. Feedback to instructors and training center officers for improvement of classroom management;

3. Data on trainee classroom performance to on-the-job instructors and group managers for the improvement of on-the-job training. (Department of the Treasury, Internal Revenue Service, Manual Supplement 04G-65, March 1977)

The manual item banks now in existence are being converted to computerized banks with plans for installing computer-assisted instruction terminals in Regional Training Centers. Test results are to be used solely for improving the effectiveness of classroom instruction.

The Social Security Administration has also found criterion-referenced testing principles and tests to be well-suited to its training needs. It has developed an extensive programmed-instruction system based on criterion-referenced testing principles for use with on-the-job training programs. The following conditions were established for selecting materials for inclusion in this program: subject matter must be of relatively stable content; mastery criterion of the subject matter should be readily identifiable; subject matter must be "must learn" types of information.

Following selection of training materials, the program objectives

must be in accord with the following criteria:

--must describe what, under what conditions, and to what degree trainees should be able to perform at the end of instruction;

--must be an overt action, i.e., one that can be observed and measured;

--the behavior described in objectives must be identical to or a representative sample of the terminal behavior that the trained person is to perform on the job;

--conditions imposed during training should be identical to those that are imposed on the employee on the job;

--performance standards must provide precise means by which the trainee's accomplishments can be measured and must state clearly and specifically what is to be required;

--minimal acceptable performance on the criterion test should be equal to the performance standards required by the objectives. (Social Security Administration, Office of Management and Administration, OHR/DTCD Pub. No.'s 90, 94, 95, 113, 115, 116, 117)

The Social Security Administration emphasizes that the CRT items developed to evaluate the trainees and the training programs are the most important elements in the process. Care must be taken to make sure that the criterion-referenced items require the trainees to perform the behavior(s) identified in the objectives and that they are controlled by the behaviors, conditions, and performance standards required by the objectives. It is also essential to ensure that the items are completely compatible with the behavior and condition elements of the related program objectives.

The Civil Service Commission's Bureau of Training turned, in the mid-70's, to a consideration of criterion-referenced testing principles to evaluate training programs. The primary goals were to: justify training on the basis of actual accomplishments; improve training courses; and facilitate

individual learning. It was expected that CRTs with their emphasis on individual accomplishment relative to established standards and content valid job-related tasks and activities would best facilitate the accomplishment of the above goals.

The Bureau of Training took the stance that properly developed and conducted CRTs when used as an integral part of the training program can:

1. force trainees to confront their need for training;
2. give trainees the feedback and challenge necessary for active involvement in their training;
3. motivate trainees to strive harder;
4. induce a favorable attitude toward training and self by demonstrating to the trainee his/her growing proficiency;
5. present trainees with a conception of excellence;
6. determine underlying causes of repeated learning difficulties.

The Bureau is in the process of preparing definitive guidelines on the use of CRTs relative to training.

The latter years of the decade have seen an upsurge in the acceptance and use of CRTs and criterion-referenced testing principles in the Federal government. There are movements underway that could presage a surge in the use of CRTs in the government. One such movement is away from broadband examining and towards decentralization of testing functions. That is, the emphasis will swing from centralized control of testing to individual agency responsibility. The genesis of this movement has its roots in proposals to decentralize many of the Federal government's responsibilities, attacks on broadband examining due to findings of adverse impact, and the political goal of a representative workforce. Related to these issues is a

movement to discover alternative selection procedures which will reduce adverse impact while ensuring the selection of qualified employees. Criterion-referenced tests would appear to be particularly well-suited to these ends.

Properly developed and utilized CRTs would also meet the spirit as well as the specific requirements for content validation of selection instruments outlined in the Uniform Guidelines on Employee Selection Procedures (1978). These Uniform Guidelines stipulate the conditions for content validity as follows: an emphasis on observable work behaviors or observable work products is appropriate; the selection procedures must be a representative sample of the content of the job; and a selection procedure measuring a skill or ability should either closely approximate an observable work behavior or its product should closely approximate an observable work product. (p. 38302)

Presently evolving is a use of criterion-referenced testing principles that will have great implications and may prove to be the most far-reaching application of such principles in the Federal government. The impetus for this movement is the Civil Service Reform Act of 1978 (CSRA) which has called for the implementation of a new performance appraisal system, one that ensures that all employees have performance standards based on the requirements of their positions. The appraisal systems now in use in the government are basically norm-referenced in nature and generally call for summary adjective ratings--e.g., outstanding, satisfactory, unsatisfactory--for personal characteristics and traits such as: working relationships, adaptability, creativity, responsibility, and independence. This system has been described as a short, uncomfortable, and often irrelevant ritual to get through once a year for each employee.

The new performance appraisal system mandated by the CSRA calls for objective performance standards based on the job in question and tailored to the job in question. The performance standards must be communicated to employees at the beginning of each appraisal period. The actual job performance of each employee will be compared to previously established job performance standards for each aspect of the job on which an employee will be appraised. The performance appraisals will provide the basis for training, rewarding, reassigning, reducing in grade, retaining in reductions in force, and removing of employees. Although idealistically speaking the new system far surpasses the old, the implementation of such a system is fraught with problems and difficulties. Since personnel decisions and actions will be taken based on the new performance appraisal system, any plan must theoretically meet the requirements of the Uniform Guidelines. There is bound to be considerable resistance from supervisors comfortable with and used to the old system. There will be a great amount of work involved in analyzing jobs and deriving valid objective standards on which to appraise employees. The CSRA also specifies that each system include some critical elements where a critical element is defined as:

> a component of an employee's job that is of sufficient importance that performance below the minimum standard established by management requires remedial action and denial of a within-grade increase, and may be the basis for removing or reducing the grade level of that employee. Such action may be taken without regard to performance on other components of the job. (Office of Personnel Management Personnel Manual, p. 5)

Defining or delineating which elements of a job are critical in terms of this definition is not an easy task. Despite these problems, a properly designed, developed, and implemented new performance appraisal system should be far superior to the old system. Such a system has the potential to:

References

Department of the Treasury, Internal Revenue Service. Test item construction workshop on criterion-referenced testing, Training #9919M Module 1. Washington, D.C.: Author, 1975.

Department of the Treasury, Internal Revenue Service. Guidelines for centralized testing program, Manual Supplement 04G-65. Washington, D.C.: Author, 1977.

Gast, I. Test use in training: An issue paper. U. S. Office of Personnel Management, 1979 (unpublished).

Public Law 95-454 95th. Congress. Civil Service Reform Act of 1978. Washington, D.C.: Author, Oct. 1978.

Social Security Administration, Office of Management and Administration. Criterion test. OHR/DTCD Pub. No. 116 (12-70). Washington, D.C.: Author.

Social Security Administration, Office of Management and Administration. Introduction to programmed instruction. OHR/DTCD Pub. No. 113 (12-70). Washington, D.C.: Author.

Social Security Administration, Office of Management and Administration. Program frames. OA-TCD, Pub. No. 94-70 (1-74). Washington, D.C.: Author.

Social Security Administration, Office of Management and Administration. Program objectives. OHR/DTCD Pub. No. 115 (12-70). Washington, D.C.: Author.

Social Security Administration, Office of Management and Administration. Program outline. OHR/DTCD Pub. No. 117 (12-70). Washington, D.C.: Author.

Swezey, R.W. & Pearlstein, R.B. Developing criterion-referenced tests. (287-AR18(2)-IR-0974-RWS-RBP), Applied Science Associates, Inc., Reston International Center, Reston, Va.

U.S. Equal Employment Opportunity Commission, U.S. Civil Service Commission, U.S. Department of Labor, & U.S. Department of Justice. Adoption by four agencies of the Uniform guidelines on employee selection procedures (1978). Federal Register, 1978, 43, 38290-38315.

U.S. Office of Personnel Management. Office of Personnel Management Personnel Manual. Chapter 430, Installment 148, p. 5. Washington, D.C.: Author, 1979.